Vol. 8, No. 3, June 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Residual-Based Hybrid SARIMA–LSTM for Bali Tourism Demand Forecasting Using Google Trends

Junaedi[1], Aditiya Hermawan[2], Yusuf Kurnia[3], Ardiane Rossi Kurniawan Maranto[4]

[1,4]Information System Department, Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia

[2,3]Informatics Department, Faculty of Science and Technology, Buddhi Dharma University, Tangerang, Indonesia

Received:
October 5, 2025
Revised:
May 10, 2026
Accepted:
May 30, 2026
Published:
June 27, 2026

Corresponding Author:
Author Name*:
Junaedi
Email*:
junaedi@ubd.ac.id

DOI:
10.63158/journalisi.v8i3.1644

Abstract. Accurate tourism demand forecasting is essential for destinations characterized by strong seasonality, nonlinear fluctuations, and post-pandemic recovery uncertainty. This study develops a residual-based hybrid SARIMA–LSTM model for forecasting monthly international tourist arrivals to Bali, Indonesia, using historical arrival data and Google Trends search query data. The dataset covers January 2009 to December 2024, comprising 192 monthly observations. A chronological split was applied, with January 2009 to December 2022 used for training and January 2023 to December 2024 used for testing. SARIMA was employed to capture linear and seasonal structures, while LSTM was used to learn nonlinear residual patterns. The proposed model was compared with SARIMA, Random Forest, standalone LSTM, and SARIMA–RF using RMSE, MAPE, and R^2. The SARIMA–LSTM model achieved the best performance, with RMSE = 35,915.36, MAPE = 5.64%, and R^2 = 0.68, compared with SARIMA, which obtained RMSE = 37,052.68, MAPE = 5.70%, and R^2 = 0.65. These findings indicate that residual-based hybridisation provides incremental forecasting improvement. However, the independent contribution of Google Trends is not separately isolated in this study and should therefore be interpreted cautiously as a complementary behavioural signal within the proposed forecasting framework.

Keywords: Bali tourism demand forecasting, Google Trends, SARIMA–LSTM, Residual learning, Search Query Data, Tourism Analytics.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Tourism demand forecasting has become an essential component of evidence-based planning in destinations that rely heavily on international visitor flows. Accurate forecasts enable policymakers and tourism stakeholders to anticipate demand fluctuations, allocate infrastructure capacity, design marketing strategies, and manage risk under uncertain conditions. This need has become increasingly important after the COVID-19 pandemic, which disrupted global mobility and produced structural breaks in historical tourism patterns [1], [2]. Bali, Indonesia, provides a particularly relevant empirical context because it combines strong seasonal demand, high dependence on international tourism, and substantial post-pandemic recovery dynamics. These characteristics make Bali not merely a destination-specific case, but a methodologically challenging setting for evaluating forecasting models that must capture both recurring seasonal patterns and irregular nonlinear fluctuations [3], [4].

Classical statistical forecasting models remain widely used in tourism analytics because of their interpretability and ability to represent temporal structures. Among these models, Seasonal Autoregressive Integrated Moving Average (SARIMA) has been frequently applied to monthly tourism data due to its capacity to model trend, autocorrelation, and annual seasonality [5], [6], [7]. SARIMA is particularly useful when demand follows relatively stable seasonal cycles, as is common in tourism destinations affected by holiday periods, school vacations, and recurring cultural events. However, despite its robustness as a baseline method, SARIMA relies on linear assumptions and may not adequately capture nonlinear variations caused by behavioural changes, economic shocks, digital information flows, or sudden disruptions in travel mobility [1], [2]. This limitation becomes more visible in post-crisis tourism recovery, where historical seasonal structures remain important but are no longer sufficient to fully explain demand volatility.

To overcome the limitations of purely statistical approaches, machine-learning and deep learning models have increasingly been adopted in tourism demand forecasting. Ensemble methods such as Random Forest can capture nonlinear relationships and interactions among predictors without imposing strict distributional assumptions [8], [9]. Deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, are

also widely used because they can learn temporal dependencies in sequential data [10], [11]. Nevertheless, the empirical performance of machine-learning models in tourism forecasting remains inconsistent. While several studies show that LSTM and ensemble models can improve forecasting accuracy, other findings indicate that these models may underperform when datasets are short, highly seasonal, or affected by structural disruptions [12]. This is a critical issue in tourism forecasting because monthly destination-level data often contain limited observations, making deep learning models vulnerable to overfitting and unstable generalisation.

Recent developments in time-series forecasting have introduced more advanced architectures such as Temporal Fusion Transformer, Transformer-based models, and decomposition-based approaches. These models are designed to capture long-range dependencies and complex temporal structures more effectively than traditional recurrent networks [13], [14]. However, their application in tourism forecasting remains constrained by data availability, computational complexity, and interpretability requirements. In practical destination-level forecasting, simpler hybrid models may remain preferable when the objective is not only to maximise predictive accuracy but also to maintain methodological transparency for policymakers. Therefore, the selection of forecasting models must balance predictive performance, interpretability, data requirements, and operational usability.

Hybrid forecasting models have emerged as a promising solution because they combine the strengths of statistical and machine-learning paradigms. The theoretical foundation of hybrid forecasting can be traced to [15], who argued that time-series data often contain both linear and nonlinear components that should be modelled separately [15]. In this logic, SARIMA can first be used to capture linear and seasonal structures, while machine-learning models can subsequently learn residual nonlinear patterns. Prior studies have demonstrated that hybrid models combining statistical decomposition and machine learning can outperform standalone models in several forecasting domains [10], [16]. However, the literature still shows important limitations. Many studies compare statistical and machine-learning models as competing alternatives, but fewer studies systematically evaluate residual-based hybridisation in highly seasonal tourism time series affected by structural disruption. In addition, the robustness of hybrid models under post-crisis recovery conditions remains insufficiently discussed.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Another relevant development in tourism demand forecasting is the use of exogenous behavioural indicators, including search query data. Google Trends has been considered a potential proxy for travel interest because destination-related online searches may occur before actual tourist arrivals [17]. In this study, Google Trends is incorporated as a complementary behavioural signal to enrich the forecasting framework, rather than as a variable whose independent predictive contribution is examined separately. This distinction is important because the usefulness of search query data may vary depending on keyword selection, temporal alignment, search behaviour patterns, and the strength of autoregressive and seasonal structures in the historical arrival series [18], [19]. Therefore, Google Trends is treated cautiously as an additional behavioural input within the proposed hybrid model, while the main analytical focus remains on residual-based hybridisation between SARIMA and LSTM.

Taken together, prior tourism forecasting studies indicate two central research gaps. First, residual-based hybrid forecasting remains insufficiently examined in destination-level tourism series characterized by strong seasonality and post-disruption recovery dynamics. Second, limited attention has been given to whether hybrid models provide substantial or only incremental improvements over strong seasonal statistical baselines. To address these gaps, this study proposes a residual-based hybrid SARIMA–LSTM framework for forecasting monthly international tourist arrivals to Bali. The main contribution of this study lies in evaluating whether nonlinear residual learning can improve the forecasting performance of a seasonal SARIMA baseline. Google Trends is included as a complementary behavioural signal in the modelling framework; however, this study does not claim to isolate its independent predictive effect. Accordingly, the study contributes to tourism analytics by offering a transparent and interpretable hybrid forecasting framework for highly seasonal post-disruption tourism demand.

2. METHODS

To enhance methodological transparency and reproducibility, this study follows a structured residual-based forecasting workflow. The research process consists of nine sequential stages: (1) collecting monthly international tourist-arrival data, (2) constructing the Google Trends digital behavioural signal, (3) aligning all variables into a monthly time-series format, (4) generating lagged predictors for the machine-learning component, (5)

splitting the dataset chronologically into training and testing periods, (6) estimating the SARIMA model to capture linear and seasonal structures, (7) extracting residuals from the SARIMA model, (8) training machine-learning models to predict residual components, and (9) reconstructing the final hybrid forecast by combining SARIMA predictions with predicted residuals. The performance of all models was then evaluated using RMSE, MAPE, and R^2.

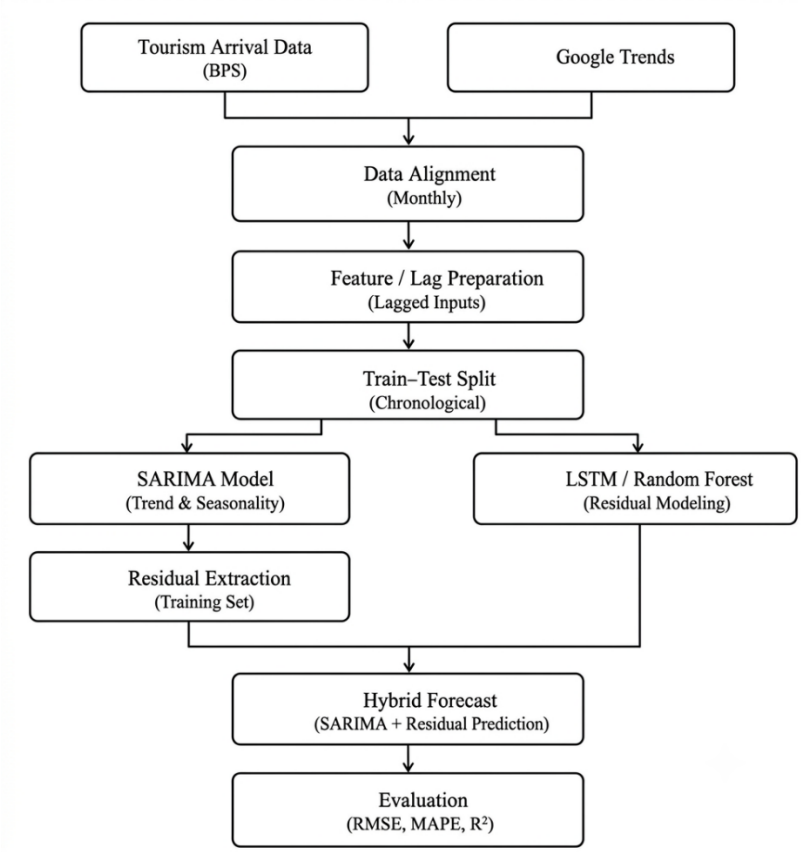

Figure 1. Research Design

Figure 1 presents the operational workflow of the proposed residual-based hybrid forecasting framework. The process begins with the collection of monthly international tourist-arrival data from Statistics Indonesia (BPS). A Google Trends series is then constructed as a digital behavioural signal representing online travel interest related to Bali. Both data sources are converted into monthly observations and aligned by calendar month to ensure temporal consistency.

After preprocessing, lagged variables are generated from historical tourist arrivals and Google Trends values. The dataset is then divided chronologically into a training period

from January 2009 to December 2022 and a testing period from January 2023 to December 2024. SARIMA is first estimated using the training data to model linear and seasonal components. The residuals produced by SARIMA are then used as the target variable for machine-learning residual prediction. Finally, the hybrid forecast is reconstructed by adding the SARIMA forecast and the predicted residual component. This workflow ensures that the statistical and machine-learning components are connected sequentially and that no information from the testing period is used during model training.

2.1. Data Description

This study uses monthly time-series data on international tourist arrivals to Bali obtained from the Central Statistics Agency (BPS) of Indonesia. The dataset covers January 2009 to December 2024, consisting of 192 observations, and includes the COVID-19 period, allowing the model to capture long-term trends, seasonal patterns, and structural disruptions in tourism demand. To enrich the forecasting framework, Google Trends data are incorporated as a digital behavioural signal representing online travel interest related to Bali.

Table 1 summarises the operational definition of the variables used in this study. International tourist arrivals serve as the dependent variable, measured monthly in number of visitors, while the Google Trends index is used as an exogenous predictor normalised on a 0–100 scale. The dataset is divided chronologically into a training set from January 2009 to December 2022 (N = 168) and a testing set from January 2023 to December 2024 (N = 24), preserving temporal order and enabling evaluation on unseen future observations.

Table 1. Variable Description

Variable	Type	Operational Definition	Source	Unit	Temporal Resolution
International Tourist Arrivals	Dependent	Total number of monthly international tourist arrivals to Bali	BPS (Statistics Indonesia)	Persons	Monthly

Variable	Type	Operational Definition	Source	Unit	Temporal Resolution
Search Trends Index	Independent (Exogenous)	Google Trends index representing global search interest for tourism-related keywords (e.g., "Bali travel"), normalised between 0–100	Google Trends	Index (0–100)	Monthly

2.2. Google Trends Signal

Google Trends was used to construct a digital behavioural signal representing online travel interest related to Bali. The final search query used in this study was "Bali travel", selected because it directly reflects international travel-related search behaviour associated with the destination. The geographic scope was set to worldwide to align the search signal with international tourist arrivals rather than domestic travel intention. The Google Trends category was restricted to Travel to reduce noise from unrelated searches involving Bali as a cultural, residential, or general-interest term.

The Google Trends index was downloaded for the period January 2009 to December 2024 and transformed into monthly observations. Because Google Trends reports relative search interest on a scale from 0 to 100, the resulting series represents normalized search intensity rather than absolute search volume. The Google Trends data were aligned with monthly tourist-arrival data using the same calendar month. To account for the possibility that online search behaviour precedes actual travel, contemporaneous Google Trends values and lagged versions at one, two, and three months were constructed. These lag structures were selected a priori based on the assumption that international travel planning and destination-related online searches may occur several months before actual arrival. The lagged Google Trends variables were used only within the machine-learning component as complementary behavioural predictors, while SARIMA was estimated using the original tourist-arrival time series.

2.3. Data Preprocessing

Prior to model development, the dataset underwent preprocessing to ensure consistency and suitability for time-series modelling. Monthly international tourist-arrival data and Google Trends data were aligned using calendar-month timestamps. Missing values in the Google Trends series, which accounted for less than 2% of observations, were handled using linear interpolation to preserve temporal continuity [20].

Lagged variables were constructed from historical tourist arrivals and the Google Trends index. The arrival series was transformed into 12-month lookback windows to represent annual seasonality, while Google Trends was represented using contemporaneous and lagged values of one to three months. These lag choices were based on the assumption that travel-related online search behaviour may occur before actual tourist arrivals.

To prevent data leakage, normalization parameters were estimated only from the training set and subsequently applied to the testing set. Min-Max scaling was used for variables entering the machine-learning models, while SARIMA was estimated on the original time-series structure. The dataset was split chronologically into a training set from January 2009 to December 2022 (N = 168) and a testing set from January 2023 to December 2024 (N = 24). This split preserves temporal order and evaluates forecasting performance on a post-pandemic recovery period.

2.4. Model Specification

This study employs three modelling approaches statistical, machine learning, and hybrid to capture the complex temporal dynamics of tourism demand. Each model is specified explicitly to ensure reproducibility and fair comparison.

1) SARIMA Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used as the baseline statistical approach to capture linear trends and seasonal structures in the time series. The general specification is defined as SARIMA(p, d, q)(P, D, Q)$_s$, where seasonal periodicity is set to s = 12 to reflect monthly tourism cycles. Model parameters were selected using a grid search procedure based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The optimal configuration identified is (1):

$$SARIMA(0,1,2)(0,1,1)_{12} \quad (1)$$

This specification effectively captures both short-term dependencies and annual seasonal patterns, consistent with prior tourism forecasting studies [9], [10].

2) LSTM Model

The Long Short-Term Memory (LSTM) network is employed to model nonlinear temporal dependencies that cannot be captured by statistical models. The architecture is designed to balance predictive performance and computational efficiency, given the relatively limited size of tourism time-series data.

Table 2. LSTM Configuration

Parameter	Value
Input window (lookback)	12 months
Number of layers	1
Units per layer	50
Activation function	tanh
Optimizer	Adam
Learning rate	0.001
Batch size	16
Epochs	100
Early stopping	Yes (patience = 10)

The lookback window of 12 months is selected to capture annual seasonality (see in Table 2). Early stopping is applied to prevent overfitting, which is a common issue in small time-series datasets [10], [11].

3) Random Forest Model

Random Forest (RF) was employed as a non-parametric ensemble learning method to capture nonlinear relationships between lagged tourist arrivals, lagged Google Trends values, and the residual structure generated by the SARIMA model [21], [22]. Unlike linear time-series models, RF does not impose distributional assumptions, making it suitable for modelling irregular fluctuations and structural variability in tourism demand.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 3. Random Forest Configuration

Parameter	Value
n_estimators	100
max_depth	None
min_samples_leaf	1
Criterion	Mean Squared Error
Hyperparameter tuning	Default (scikit-learn)

Table 3 presents the configuration of the Random Forest (RF) model used in this study. The model employs 100 decision trees (n_estimators = 100) with no restriction on tree depth (max_depth = None), allowing it to capture complex nonlinear relationships. A minimum leaf size of one (min_samples_leaf = 1) is applied to maintain model flexibility. The Mean Squared Error (MSE) criterion is used for node splitting, ensuring consistency with the evaluation metrics.

RF operates using bootstrap aggregation, where each tree is trained on randomly sampled data and feature subsets, reducing variance and improving generalisation. Lag-based features derived from historical tourist arrivals are used as inputs, enabling the model to capture temporal dependencies implicitly. Default hyperparameters from the scikit-learn library are adopted to provide a robust baseline configuration [8].

4) Residual-Based Hybrid Model

The proposed hybrid framework integrates SARIMA and machine-learning models through a residual learning mechanism. This approach follows the theoretical foundation proposed by [15], in which linear and nonlinear components of a time series are modelled separately to improve forecasting performance. In this study, the SARIMA model is first fitted to the training data to generate baseline forecasts that capture the linear trend and seasonal structure of tourist arrivals. The residual series is then computed as the difference between the observed values and the SARIMA predictions. These residuals represent the nonlinear variation that is not explained by the statistical model. Subsequently, machine-learning models, namely LSTM and Random Forest, are trained using the residual series derived exclusively from the training set. The trained models are then used to predict residual values in the testing period. Finally, the hybrid forecast

is reconstructed by adding the SARIMA forecast to the predicted residual component generated by the machine-learning model. Formally, the final hybrid prediction can be expressed as follows:

$$\hat{Y}_t^{Hybrid} = \hat{Y}_t^{SARIMA} + \hat{e}_t^{ML} \quad (2)$$

Equation (2) represents the final hybrid forecast and $\hat{e}_t^{ML}$ represents the residual predicted by the machine-learning model, either LSTM or Random Forest. This formulation ensures that SARIMA models the linear and seasonal structure, while the machine-learning component focuses only on the remaining nonlinear residual variation.

To ensure methodological rigor, residual learning is performed strictly on the training set, and no information from the testing period is used during model training. Exogenous variables are incorporated only into the machine-learning component, allowing SARIMA to focus on modelling deterministic seasonal structure while the machine-learning model captures nonlinear relationships and behavioural fluctuations. This separation of modelling processes reduces model bias and improves forecasting robustness [15].

2.5. Validation and Implementation

This study used a chronological train–test split to evaluate forecasting performance on unseen future observations while preserving the temporal ordering of the time series. The training period covered January 2009 to December 2022, while the testing period covered January 2023 to December 2024. This testing window was selected because it represents a post-pandemic recovery period in which tourism demand became more volatile and policy-relevant.

The chronological split was used as an initial validation strategy. Rolling-origin or expanding-window validation was not applied in the main experiment because the study used a single-destination monthly dataset with only 192 observations. Applying multiple rolling windows would reduce the effective training length in several folds, particularly for LSTM-based models that require sufficient sequential observations. Therefore, the single train–test split provides an initial assessment of forecasting performance, while temporal robustness across multiple forecasting origins is recommended for future research. Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). RMSE was used

to measure the absolute magnitude of forecast errors, MAPE was used to assess relative forecasting accuracy, and R^2 was used to indicate the proportion of variance explained by the model during the testing period.

All experiments were implemented in Python. SARIMA estimation was conducted using the statsmodels library. Random Forest modelling, Min-Max scaling, and performance evaluation were conducted using scikit-learn. The LSTM model was implemented using TensorFlow/Keras. Data manipulation and numerical computation were performed using pandas and NumPy. Normalization parameters were estimated only from the training set and then applied to the testing set to prevent data leakage.

3. RESULTS AND DISCUSSION

3.1. Stationarity Assessment and SARIMA Identification

Before model estimation, the monthly international tourist arrivals series was examined for stationarity using the Augmented Dickey–Fuller (ADF) test. The initial test indicated that the original series was non-stationary, with a p-value of 0.0885, exceeding the 0.05 significance threshold. This finding suggests the presence of trend and seasonal components in the raw series, which is consistent with the long-term cyclical behaviour typically observed in tourism demand data.

To address this issue, first-order differencing was applied. The repeated ADF test on the differenced series produced a p-value of 0.0411, indicating that the transformed series had become stationary and was therefore suitable for SARIMA modelling. This result confirms that differencing successfully removed the non-stationary component while preserving the temporal structure necessary for subsequent forecasting.

The stationarity assessment was followed by inspection of the autocorrelation function (ACF) and partial autocorrelation function (PACF), as presented in Figure 2. The ACF and PACF plots were used to guide the identification of autoregressive and moving average terms, while final model selection was based on a grid-search procedure evaluated through Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Among the candidate specifications, SARIMA$(0,1,2)(0,1,1)_{12}$ achieved the lowest AIC and BIC values, indicating the best trade-off between model fit and parsimony. This result

confirms that the tourist arrivals series exhibits both short-term dependence and strong annual seasonality, which justify the use of a seasonal statistical model as the baseline component of the hybrid framework.

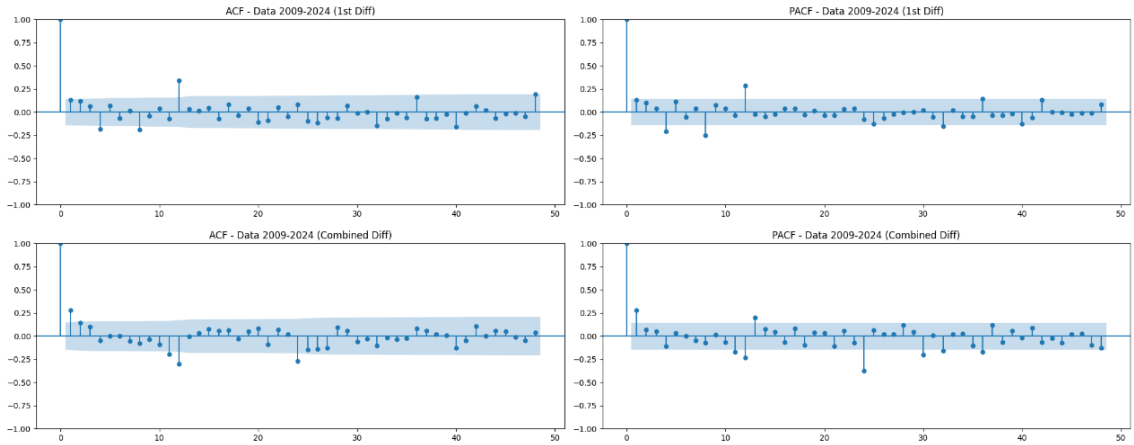

Figure 2. ACF and PACF Plots

3.2. Residual Diagnostic Analysis

Residual diagnostics were conducted to evaluate whether the SARIMA model had adequately captured the main temporal structure of the series. As shown in Figure 3, the time-series residual plot indicates that the residuals fluctuate randomly around zero, with no obvious systematic trend remaining. This suggests that the baseline SARIMA model successfully represents the dominant linear and seasonal structure of the series. The histogram and kernel density estimate further show that the residual distribution is centred around zero and approximately symmetric, although a number of extreme residuals remain visible in the tails. These deviations indicate that the model does not fully explain all irregular movements in the data, especially during periods of abrupt demand changes. This is analytically important because it provides justification for the residual-learning stage of the hybrid framework: if the residuals still contain nonlinear structure, then a machine-learning model may add predictive value by learning what SARIMA leaves unexplained.

The residual ACF also indicates that most autocorrelation coefficients fall within the confidence bounds, implying that substantial serial dependence has been removed. Nevertheless, the existence of large residual deviations during selected periods suggests that the remaining errors are not purely random in an economic sense, particularly during

post-disruption recovery phases. In other words, while the SARIMA model is statistically adequate as a seasonal baseline, it still leaves room for improvement in modelling nonlinear volatility and irregular recovery patterns.

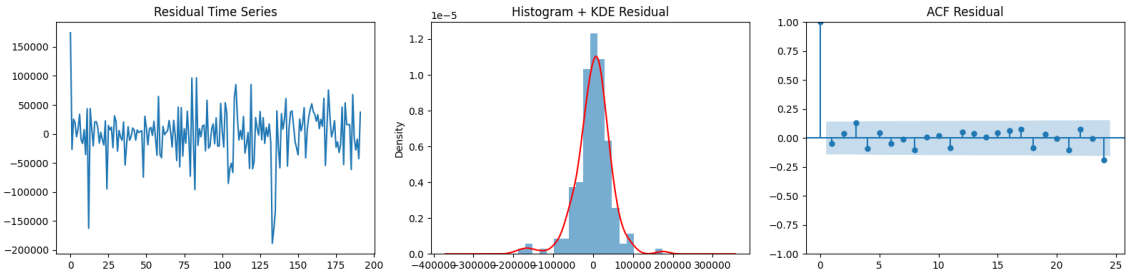

Figure 3. Residual diagnostics of the baseline SARIMA model

3.3. Comparative Forecasting Performance

Table 4 presents the comparative forecasting performance of the evaluated models based on RMSE, MAPE, and R^2. The SARIMA–LSTM model achieved the best overall performance, with the lowest RMSE (35,915.36), the lowest MAPE (5.64%), and the highest R^2 (0.68). Compared with the SARIMA baseline, SARIMA–LSTM reduced RMSE by approximately 3.07%, reduced MAPE by approximately 1.05%, and increased R^2 by approximately 4.62%. These results indicate that the proposed residual-based hybrid model provides an incremental improvement over the seasonal statistical baseline.

However, the comparison also shows that hybridisation and machine-learning modelling did not uniformly improve performance. Random Forest produced a 2.20% higher RMSE and a 3.51% higher MAPE than SARIMA, while its R^2 decreased by approximately 3.08%. The standalone LSTM model performed substantially worse than SARIMA, with RMSE increasing by approximately 42.36%, MAPE increasing by approximately 34.74%, and R^2 decreasing by approximately 53.85%. These results suggest that a standalone deep-learning model is less suitable for this dataset, most likely because the monthly destination-level series contains limited observations and strong seasonal patterns. SARIMA–RF produced a very small RMSE reduction of approximately 0.19% relative to SARIMA, but its MAPE was 2.98% higher, indicating that the improvement was not consistent across evaluation metrics. Overall, SARIMA–LSTM was the strongest model in the 24-month testing window from January 2023 to December 2024. Nevertheless, the performance gain should be interpreted cautiously because the evaluation is based on a

single chronological test period. The results support the usefulness of residual-based hybridisation, but they do not establish that the model would necessarily maintain the same superiority across multiple rolling-origin forecasting windows.

Table 4. Comparative Forecasting Performance of Baseline and Hybrid Models

Model	RMSE	MAPE	R^2
SARIMA	37,052.68	5.70%	0.65
Random Forest	37,869.65	5.90%	0.63
LSTM	52,748.91	7.68%	0.30
SARIMA–RF	36,982.60	5.87%	0.66
SARIMA–LSTM	35,915.36	5.64%	0.68

Table 5. Relative Performance Change Compared with the SARIMA Baseline

Model	RMSE Change vs SARIMA	MAPE Change vs SARIMA	R^2 Change vs SARIMA	Interpretation
Random Forest	-2.20%	-3.51%	-3.08%	Worse than SARIMA
LSTM	-42.36%	-34.74%	-53.85%	Substantially worse
SARIMA–RF	+0.19%	-2.98%	+1.54%	Mixed result
SARIMA–LSTM	+3.07%	+1.05%	+4.62%	Best overall

As shown in Table 5, SARIMA–LSTM provides the most consistent improvement over the SARIMA baseline, with better performance across all three metrics. In contrast, SARIMA–RF shows only a marginal RMSE improvement but performs worse in terms of MAPE, while standalone Random Forest and LSTM underperform compared with SARIMA.

The results also show that hybridisation did not uniformly improve all model configurations. SARIMA–RF produced a slightly lower RMSE than SARIMA, but its MAPE was higher than that of SARIMA. This suggests that the benefit of residual-based hybridisation depends on the residual learner used. In this dataset, LSTM was more effective than Random Forest in modelling the remaining residual structure after

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

SARIMA, whereas the standalone LSTM model performed poorly, likely because monthly destination-level tourism data provide a relatively small sample for deep learning models. Therefore, the comparative evidence supports SARIMA–LSTM as the strongest configuration, but the improvement should be interpreted as incremental rather than substantial.

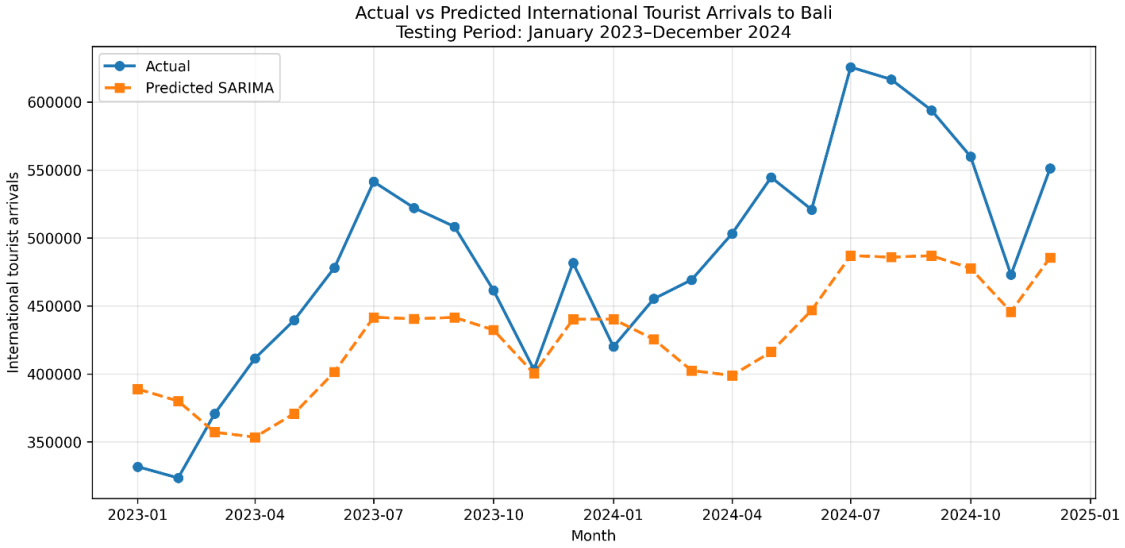

Figure 4. Comparison of actual and predicted international tourist arrivals during the testing period (January 2023–December 2024)

Figure 4 compares the actual and predicted international tourist arrivals during the testing period from January 2023 to December 2024. The visual pattern indicates that SARIMA was able to follow the broad seasonal movement of the series, but deviations remained during periods of sharper fluctuation. The hybrid SARIMA–LSTM model appears to provide a closer adjustment to the actual series in phases where the SARIMA forecast leaves residual variation, particularly when the movement of tourist arrivals changes more rapidly. This suggests that the LSTM residual-learning component contributes by correcting part of the remaining nonlinear error after the seasonal structure has been modelled by SARIMA. Nevertheless, the visual improvement should be interpreted alongside the quantitative results in Table 4. The reduction in RMSE and MAPE is measurable but modest, indicating that most of the predictable structure is already captured by the SARIMA baseline. Therefore, Figure 4 supports the interpretation that SARIMA–LSTM improves forecast accuracy incrementally rather than producing a major transformation in forecasting performance. The evidence should also be understood

within the limitation that the visual comparison is based on one 24-month post-pandemic testing window.

.

3.4. Assessment of Digital Behavioural Signal Contribution

Google Trends was incorporated in this study as a digital behavioural signal representing online search interest related to Bali tourism. However, the present results do not separately isolate the independent predictive contribution of Google Trends. Therefore, the comparative results in Table 4 should not be interpreted as direct evidence that Google Trends independently improves forecasting accuracy. Rather, the results primarily demonstrate the performance of the overall residual-based hybrid framework, particularly the combination of SARIMA for modelling seasonal structure and LSTM for modelling remaining residual variation.

This distinction is important because the performance gain of SARIMA–LSTM may be associated with the residual-learning mechanism, the inclusion of search-based behavioural information, or the interaction between both components. Since this study does not provide a separate model comparison that isolates Google Trends from historical arrival lags, the role of Google Trends should be interpreted cautiously. In this study, Google Trends is best understood as a complementary behavioural input within the hybrid forecasting framework rather than as a demonstrated driver of forecasting improvement. Accordingly, the manuscript avoids claiming that Google Trends directly improves forecasting responsiveness. The more appropriate interpretation is that Google Trends provides theoretically relevant search query information that may enrich the residual-learning component, while its independent predictive value should be examined more rigorously in future studies using controlled model comparisons.

3.5. Discussion

The results demonstrate that hybrid models, particularly the SARIMA–LSTM configuration, achieve the best predictive performance across all evaluated metrics. However, the magnitude of improvement over the baseline SARIMA model is relatively modest (approximately 3% reduction in RMSE). This suggests that while residual-based hybridization enhances forecasting accuracy, the gain should be interpreted cautiously rather than as a substantial breakthrough. Similar findings have been reported in recent

studies, where hybrid approaches provide consistent but incremental improvements over statistical models [23].

From a comparative perspective, the observed improvement is smaller than that reported in some recent tourism forecasting studies. For instance, [2] reported RMSE reductions of approximately 12% using advanced hybrid neural architectures. This discrepancy may be attributed to differences in dataset size, feature richness, and modelling complexity. In contrast, the present study operates on a relatively constrained dataset with strong seasonal patterns, where statistical models already capture a large portion of the variance. As a result, the marginal contribution of nonlinear modelling becomes more limited, leading to smaller but still consistent gains.

The relatively weak performance of the standalone LSTM model further reinforces this interpretation. Despite its theoretical capability to model long-term dependencies, LSTM underperforms significantly in this study, which is consistent with prior findings in tourism forecasting contexts characterised by limited data and strong seasonality [11], [12]. Deep learning models typically require large datasets to generalise effectively, and their performance may degrade when applied to short, noisy, or structurally unstable time series. This highlights the importance of explicitly modelling deterministic seasonal structures, which SARIMA captures more effectively.

Conversely, the performance of Random Forest remains comparable to SARIMA but does not provide significant improvements. This aligns with findings by [24], [25], who reported that ensemble methods such as Random Forest are more effective in handling disrupted or highly irregular datasets, particularly during periods such as the COVID-19 pandemic. In the present case, however, the dominance of strong seasonal patterns appears to limit the advantage of purely nonlinear models, further supporting the rationale for hybridization.

The effectiveness of the residual-based hybrid approach can be explained by its ability to decompose the forecasting problem into complementary components. The SARIMA model captures deterministic linear and seasonal structures, while the machine-learning component models the remaining nonlinear residuals. This separation aligns with the theoretical framework proposed by [15] and has been widely supported in cross-domain

forecasting literature [10]. By modelling different aspects of the data independently, the hybrid approach reduces bias and improves robustness, particularly in datasets characterised by heterogeneous temporal dynamics.

Despite these advantages, several methodological limitations should be acknowledged. First, the study relies on a relatively simple hybrid framework with limited exogenous variables, which may not fully capture the complexity of tourism demand drivers. Second, the use of a single chronological split, consisting of 168 training observations and 24 testing observations, may limit the robustness of performance evaluation. Although this split preserves temporal ordering and enables evaluation on a post-pandemic testing window, it does not fully capture how model performance may vary under different forecasting origins. Rolling-origin or expanding-window validation would provide stronger evidence of temporal generalisation, but such validation was not implemented in the present study because the dataset consists of a single-destination monthly series with a limited number of observations. Therefore, the reported performance should be interpreted as an initial out-of-sample evaluation rather than conclusive evidence of generalisable superiority. From a practical perspective, even modest improvements in forecasting accuracy can have meaningful implications for tourism management. More accurate demand predictions enable policymakers to better anticipate fluctuations in visitor arrivals, optimise resource allocation, and design targeted marketing strategies. In the context of Bali, where tourism plays a critical role in the regional economy, improved forecasting can support infrastructure planning, mitigate overcrowding, and enhance sustainability initiatives. For example, more reliable forecasts can inform capacity management strategies during peak seasons and guide recovery planning in post-disruption periods.

The limited performance gain of SARIMA–LSTM can be explained by three factors. First, the monthly tourist-arrival series contains strong annual seasonality, which SARIMA is already well suited to capture. When the statistical baseline explains a large share of the systematic temporal structure, the residual component available for machine-learning correction becomes smaller. Second, the testing period contains only 24 monthly observations, which limits the stability of aggregate performance metrics and reduces the statistical power of forecast comparison tests. Third, Google Trends may not fully capture actual travel realisation because search behaviour can be affected by

information-seeking, trip planning, media exposure, and non-travel interest in Bali. These factors explain why the hybrid model improves forecast accuracy only modestly rather than producing a large performance gain. The role of Google Trends should therefore be interpreted carefully. Theoretically, search behaviour may serve as a leading indicator of tourism demand because potential visitors often search for destination information before travelling. However, in this study, the independent predictive contribution of Google Trends is not separately isolated from the residual-learning mechanism. Therefore, Google Trends should be understood as a supplementary behavioural signal within the proposed hybrid framework rather than as a primary determinant of forecasting accuracy. This interpretation is consistent with the strong seasonal structure of the Bali arrival series, where historical arrivals remain the dominant source of predictive information.

4. CONCLUSION

This study examined a residual-based hybrid SARIMA–LSTM framework for forecasting monthly international tourist arrivals to Bali. The results show that SARIMA–LSTM achieved the best overall performance among the evaluated models, with RMSE = 35,915.36, MAPE = 5.64%, and R^2 = 0.68. Compared with the SARIMA baseline, the improvement was modest, indicating that SARIMA already captured much of the linear and seasonal structure of the tourist-arrival series. Therefore, the main demonstrated contribution of this study is the use of residual-based hybridisation, where LSTM provides incremental improvement by modelling the nonlinear residual variation left by SARIMA. Although Google Trends was included as a search-based behavioural input, its marginal predictive effect remains unverified because the study did not separately isolate its independent contribution. Thus, Google Trends should be interpreted as an exploratory complementary signal rather than as a proven driver of forecasting improvement. Future research should conduct controlled with-vs-without Google Trends comparisons, apply rolling-origin or expanding-window validation, use multi-destination datasets, and incorporate additional predictors such as economic indicators, flight capacity, hotel occupancy, and policy-related variables.

ACKNOWLEDGMENT

The authors gratefully acknowledge Buddhi Dharma University for valuable institutional support throughout the completion of this research. Their support has contributed significantly to facilitating the research process and the development of this study.

REFERENCES

[1] S. Gricar, "Tourism Forecasting of ' Unpredictable ' Future Shocks : A Literature Review by the PRISMA Model," vol. 16, no. 12, pp. 1-13, 2023, doi: 10.3390/jrfm16120493.

[2] Y. Zhang and W. H. Tan, "Tourism Demand Forecasting Based on a Hybrid Temporal Neural Network Model for Sustainable Tourism," vol. 17, no. 5, pp. 1–15, 2025, doi: 10.3390/su17052210.

[3] D. P. Ramadhani, A. Alamsyah, M. Y. Febrianta, L. Zulfa, and A. Damayanti, "Exploring Tourists ' Behavioral Patterns in Bali ' s Top-Rated Destinations : Perception and Mobility," vol. 19, no. 2, pp. 743–773, 2024, doi: 10.3390/jtaer19020040.

[4] I. G. B. R. Utama *et al.*, "Exploration Of The Advantages Of Tourism Branding In Bali , Indonesia," *Int. J. Prof. Bus. Rev.*, vol. 8, no. 3, pp. 1–17, 2023, doi: 10.26668/businessreview/2023.v8i3.1609.

[5] S. Sitara, W. Fatima, and A. Rahimi, "A Review of Time-Series Forecasting Algorithms for Industrial," vol. 12, no. 6, pp. 1-30, 2024, doi: 10.3390/machines12060380.

[6] D. G. Guminta, "Comparison of ARIMA and SARIMA Methods for Non-Oil and Gas Export Forecasting in East Java," *J. Apl. Sains Data*, vol. 01, no. 1, 2025, pp. 1–9, 2025, doi: 10.33005/jasid.v1i1.2.

[7] D. Nurhasanah, A. Maulidya, and M. Dwi, "Forecasting International Tourist Arrivals in Indonesia Using SARIMA Model," vol. 2, no. 6, pp. 19-25, 2022, doi: 10.20885/enthusiastic.vol2.iss1.art3.

[8] R. K. Mishra, S. Urolagin, J. A. A. Jothi, N. Nawaz, and H. Ramkissoon, "Machine Learning based Forecasting Systems for Worldwide International Tourists Arrival," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 55–64, 2021, doi: 10.14569/IJACSA.2021.0121107.

[9] D. T. Andariesta and M. Wasesa, "Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic : a multisource Internet data approach," vol. 12, no. 1, pp. 91-107, 2026, doi: 10.1108/JTF-10-2021-0239.

[10] L. Peng, L. Wang, X. Y. Ai, and Y. R. Zeng, "Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory," *Cognit. Comput.*, vol. 13, no. 1, pp. 125–138, 2021, doi: 10.1007/s12559-020-09747-z.

[11] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of deep learning for time series forecasting : architectural diversity and open challenges," *Artif. Intell. Rev.*, pp 1-79, 2025, doi: 10.48550/arXiv.2411.05793.

[12] N. M. De Jesus and B. R. Samonte, "AI in Tourism: Leveraging Machine Learning in Predicting Tourist Arrivals in Philippines using Artificial Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 816–823, 2023, doi: 10.14569/IJACSA.2023.0140393.

[13] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[14] E. U. Capoglu and A. Taherkhani, "A Comparison of Different Transformer Models for Time Series Prediction," vol. 16, no. 10, pp. 1–15, 2025, doi: 10.3390/info16100878.

[15] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003, doi: 10.1016/S0925-2312(01)00702-0.

[16] V. Arumugam and V. Natarajan, "Enhanced time series forecasting using hybrid ARIMA and machine learning models," vol. 38, no. 3, pp. 1970–1979, 2025, doi: 10.11591/ijeecs.v38.i3.pp1970-1979.

[17] P. Kaewmanee, J. Muangprathub, and W. Sae-Jie, "Forecasting tourist arrivals with keyword search using time series," *ECTI-CON 2021 - 2021 18th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. Smart Electr. Syst. Technol. Proc.*, pp. 171–174, 2021, doi: 10.1109/ECTI-CON51831.2021.9454824.

[18] Junaedi, A. H. Gunawan, V. Kuswanto, and Jonathan, "Eksplorasi Algoritma Support Vector Machine untuk Analisis Sentimen Destinasi Wisata di Indonesia," *bit-Tech*, vol. 7, no. 2, 2024, doi: 10.32877/bt.v7i2.1810.

[19] E. Christou and A. Giannopoulos, "The Evolution of Digital Tourism Marketing : From Hashtags to AI-Immersive Journeys in the Metaverse Era," vol. 17, no. 13, pp. 1–41, 2025, doi: 10.3390/su17136016.

[20] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed. Melbourne, Australia, 2021. [Online]. Available: https://otexts.com/fpp3/

[21] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[22] S. Abul and P. Sadorsky, "Machine Learning with Applications Forecasting Bitcoin price direction with random forests : How important are interest rates , inflation , and market volatility ?," *Mach. Learn. with Appl.*, vol. 9, no. 5, pp. 1-19, 2022, doi: 10.1016/j.mlwa.2022.100355.

[23] M. Milli, "Designing a residual-enhanced hybrid Prophet – LSTM framework for urban air pollution forecasting in Beijing," vol. 15, pp. 1–24, 2025, doi: 10.1038/s41598-025-27510-y.

[24] K. J. Waciko, L. A. Susanti, and R. Nur, "Forecasting Tourist Arrivals in Bali : A Grid Search-Tuned Comparative Study of Random Forest , XGBoost , and a Hybrid RF-XGBoost Model," vol. 8, no. 3, pp. 251–261, 2025, doi: 10.12962%2Fj27213862.v8i3.23334.

[25] R. Tapio and D. Tarepe, "Comparative Analysis of Random Forest and Hybrid ARIMA-random Forest Models for Student Enrollment Forecasting in Higher Education," vol. 40, no. 3, pp. 124–136, 2025, doi: 10.9734/jamcs/2025/v40i31982.